

**PRIORITY  
DOCUMENT**

SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)



REC'D 08 NOV 2000

WIPO PCT

DE 00/02940

EU

**Prioritätsbescheinigung über die Einreichung  
einer Patentanmeldung**

**Aktenzeichen:**

199 42 178.1

**Anmeldetag:**

3. September 1999

**Anmelder/Inhaber:**

Siemens AG, München/DE

**Bezeichnung:**

Verfahren zum Aufbereiten einer Datenbank für die  
automatische Sprachverarbeitung

**IPC:**

G 10 L 5/04

**Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ur-  
sprünglichen Unterlagen dieser Patentanmeldung.**

München, den 20. September 2000  
Deutsches Patent- und Markenamt  
Der Präsident  
Im Auftrag

Nietiert



## Beschreibung

Verfahren zum Aufbereiten einer Datenbank für die automatische Sprachverarbeitung

5

Die Erfindung betrifft ein Verfahren zum Aufbereiten einer Datenbank für die automatische Sprachverarbeitung, sowie ein Verfahren zum Trainieren eines neuronalen Netzwerkes zum Zuordnen von Graphemen zu Phonemen für die automatische Sprachverarbeitung und ein Verfahren zum Zuordnen von Graphemen zu Phonemen bei der Synthetisierung von Sprache bzw. bei der Erkennung von Sprache.

15

Es ist bekannt, neuronale Netzwerke für die Synthetisierung von Sprache zu verwenden, wobei die neuronalen Netzwerke einen Text, der in einer Folge von Graphemen dargestellt ist, in Phoneme umsetzen, welche von einer entsprechenden Sprachausgabevorrichtung in die korrespondierenden akustischen Laute gewandelt werden. Grapheme sind Buchstaben beziehungsweise Buchstabenkombinationen, welchen jeweils ein Laut, das Phonem, zugeordnet ist. Vor einem erstmaligen Einsatz des neuronalen Netzwerkes muß dieses trainiert werden. Dies erfolgt üblicherweise durch Verwendung einer Datenbank, die die Graphem-Phonem-Zuordnungen enthält, wodurch festgelegt ist, welchem Graphem welches Phonem zugeordnet ist.

20

Die Erstellung einer solchen Datenbank bedeutet einen erheblichen zeitlichen wie auch geistigen Aufwand, da derartige Datenbanken in der Regel nur mit Hilfe eines Sprachexperten aufgebaut werden können.

30

Der Erfindung liegt die Aufgabe zugrunde ein Verfahren zu schaffen, mit welchen auf einfache Art und Weise eine Graphem-Phonem-Zuordnungen enthaltende Datenbank erstellt werden kann.

35

gung der Phonemeinheiten und/oder Graphemeinheiten zugeordnet werden, und

5 wobei zumindest nach Schritt a) ein Korrekturschritt ausgeführt wird, mit dem Zuordnungen von Wörtern, die im Widerspruch zu den im Schritt a) ermittelten weiteren Zuordnungen stehen, gelöscht werden.

10 Erfindungsgemäß werden zunächst Wörter untersucht, die dieselbe Anzahl Grapheme und Phoneme besitzen. Die Grapheme dieser Wörter werden den Phonemen paarweise zugeordnet, wobei in einem darauffolgenden Korrekturschritt die Zuordnungen der Wörter gelöscht werden, die im Widerspruch zu den weiteren Zuordnungen stehen.

15 Mit diesem ersten Zuordnungsvorgang kann eine Großzahl der Wörter abgearbeitet werden und zudem statistisch signifikante Zuordnungen erzielt werden, die eine Überprüfung im Korrekturschritt erlauben und die auch eine Überprüfung der weiteren zu erstellenden Zuordnungen in den nachfolgenden Schritten erlauben.

20 Danach werden die Wörter untersucht, bei denen sich die Anzahl der Phoneme gegenüber der Anzahl von Graphemen unterscheidet. Bei Wörtern mit mehr Graphemen als Phonemen werden  
25 mehrere Grapheme zu Graphemeinheiten zusammengefaßt und bei Wörtern mit weniger Graphemen als Phonemen werden Phoneme zu Phonemeinheiten zusammengefaßt.

30 Nach Beendigung dieser Schritte werden die bisher nicht zuordbaren Wörter überprüft, wobei hierbei die ermittelten Phonemeinheiten und/oder die ermittelten Graphemeinheiten berücksichtigt werden.

35 Mit dem erfindungsgemäßen Verfahren wird somit schrittweise automatisch ein „Zuordnungswissen“ erstellt, das zunächst auf paarweisen Graphem-Phonem-Zuordnungen beruht und in das im

gemäßen Verfahrens ist die Erstellung einer solchen Zuordnung.

5 In Fig. 1 ist ein Ausführungsbeispiel des erfindungsgemäßen Verfahrens in einem Flußdiagramm dargestellt. In einem Schritt S1 wird das Verfahren gestartet.

10 Im Schritt S2 werden alle Wörter untersucht, die die gleiche Anzahl Grapheme und Phoneme besitzen. Die Grapheme dieser Wörter werden den entsprechenden Phonemen paarweise zugeordnet.

15 Eine derartige paarweise Zuordnung wird z.B. für das englische Wort „run“ ausgeführt, das folgendermaßen mit seinen Graphemen und Phonemen dargestellt werden kann:

Grapheme: r u n

Phoneme: r A n

20 Bei „run“ wird das Graphem „r“ dem Phonem „r“, das Graphem „u“ dem Phonem „A“ und das Graphem „n“ dem Phonem „n“ zugeordnet. Bei dieser paarweisen Zuordnung wird somit jeweils ein einziges Graphem einem einzigen Phonem zugeordnet. Dies wird für alle Wörter ausgeführt, die die gleiche Anzahl von  
25 Phoneme und Grapheme besitzen.

30 Im nachfolgenden Schritt S3 wird eine Korrektur ausgeführt, mit der die Zuordnungen der Wörter, die im Widerspruch zu dem im Schritt S2 ermittelten weiteren Zuordnungen stehen, gelöscht werden. Hierzu werden die Häufigkeiten der einzelnen Graphem-Phonem-Zuordnungen erfaßt und Graphem-Phonem-Zuordnungen die nur selten vorkommen werden gelöscht. Liegt die Häufigkeit einer bestimmten Graphem-Phonem-Zuordnung unter einem vorbestimmten Schwellwert, so werden die entsprechenden  
35 Graphem-Phonem-Zuordnungen gelöscht. Der Schwellwert liegt z.B. im Bereich von einer Häufigkeit von 10 bis 100. Je nach Umfang des Vokabulars der Ausgangsdatenbank kann der

Im Schritt S6 werden die Wörter, die mehr Grapheme als Phoneme besitzen und im Schritt S4 nicht korrekt zugeordnet werden konnten, von neuem untersucht, wobei jeweils ein einzelnes  
5 Graphem einem einzelnen Phonem in Leserichtung (von links nach rechts) zugeordnet wird. Jede einzelne Zuordnung wird überprüft, ob sie den bisher ermittelten Zuordnungen entspricht. Ergibt diese Überprüfung, daß eine Graphem-Phonem-Zuordnung nicht den bisherigen Zuordnungen entspricht, das  
10 heißt, daß sie nicht die notwendige Häufigkeit besitzt, geht das Verfahren auf die letzte Graphem-Phonem-Zuordnung zurück und verbindet das Graphem dieser Graphem-Phonem-Zuordnung mit dem in Leserichtung nächstem Graphem zu einer Graphemeinheit. Die übrigen Phoneme und Grapheme werden dann wieder einander  
15 einzeln zugeordnet, wobei wiederum jede einzelne Graphem-Phonem-Zuordnung überprüft wird.

Bei diesem Verfahrensschritt können innerhalb eines Wortes eine oder mehrere Graphemeinheiten erzeugt werden, wobei die  
20 Graphemeinheiten in der Regel zwei Grapheme umfassen. Es ist jedoch auch möglich, daß die Graphemeinheiten drei oder mehr Grapheme umfassen können.

Ein Wort, bei dem der Schritt S6 zu einer erfolgreichen Zuordnung führt ist z.B. das englische Wort „abacement“:  
25

Grapheme: a b a s e m e n t

Phoneme: x b e s m i n t

30 Bei „abacement“ erfolgt die paarweise Zuordnung korrekt bis zum Graphem „e“ das zunächst dem Phonem „m“, zugeordnet wird. Diese Zuordnung steht im Widerspruch zu den bisher ermittelten Zuordnungen, weshalb das Verfahren auf die letzte erfolgreiche Zuordnung des Graphems „s“ zum Phonem „s“ zurückgeht  
35 und das Graphem „s“ mit dem Graphem „e“ zur Graphemeinheit „se“ verbindet. Die weitere paarweise Zuordnung der Grapheme

nung verläuft korrekt bis zum Graphem „g“ das zunächst dem Phonem „G“ zugeordnet wird. Diese Zuordnung steht im Widerspruch zu den bisher ermittelten Zuordnungen. Als einziges Phonem dem kein Graphem zugeordnet werden kann, verbleibt das  
5 Phonem „G“. Diesem Phonem „G“ werden nun die restlichen Grapheme „n“ und „g“, die zu einer Graphemeinheit zusammengefaßt werden, zugeordnet.

10 Im Schritt S9 werden die im Schritt S8 untersuchten Wörter, die nicht vollständig erfolgreich zugeordnet worden sind, markiert und deren Zuordnungen werden wiederum gelöscht.

15 Im Schritt S10 werden die Wörter, die weniger Grapheme als Phoneme besitzen untersucht, wobei die einzelnen Grapheme den einzelnen Phonemen paarweise zugeordnet werden, wobei die Grapheme auch den zu den zugeordneten Phonemen benachbarten Phonemen zugeordnet werden. Von all diesen Zuordnungen wird die jeweilige Häufigkeit bestimmt, und falls festgestellt wird, daß ein Graphem zwei benachbarten Phonemen mit einer  
20 großen Häufigkeit zugeordnet werden kann, werden diese beiden Phoneme zu einer Phonemeinheit zusammengefaßt, falls die beiden Phoneme zwei Vokale oder zwei Konsonanten sind.

25 Ein Wort, bei dem der Schritt S9 zu einer korrekten Zuordnung führt, ist z.B. das englische Wort „axes“:

Grapheme: a x e s

Phoneme: @ ks i z

30 Bei „axes“ ergibt die Zuordnungen des Graphems „x“ zu den Phonemen „k“ und „s“ jeweils eine Häufigkeit, die über einem vorbestimmten Schwellwert liegt, so daß diese beiden Phoneme zur Phonemeinheit „ks“ zusammengefaßt werden. Die übrigen Grapheme und Phoneme werden wiederum paarweise zugeordnet.  
35

Grapheme: a c c u m u l a t e d

Phoneme: x k y u m y x l e t I d

5 Bei diesem Wort werden zunächst die Phoneme „y“ und „u“ bzw.  
„y“ und „x“ durch die Phonemeinheiten „yu“ bzw. „yx“ ersetzt.  
Da diese Phonemeinheiten bereits bei den vorhergehenden  
Schritten ermittelt worden sind, wird im Schritt S14 von der  
Option Gebrauch gemacht, daß auch die Graphemeinheiten be-  
10 rücksichtigt werden, so wird für die beiden Grapheme „c“ und  
„c“ die Graphemeinheit „cc“ verwendet. Die paarweise Zuord-  
nungen der einzelnen Grapheme bzw. Graphemeinheiten zu den  
einzelnen Phonemen bzw. Phonemeinheiten ergibt eine korrekte  
Zuordnung.

15

Wird von der Option der Berücksichtigung der Graphemeinheiten  
kein Gebrauch gemacht, so werden, wie es im Schritt S6 der  
Fall ist, die einzelnen Grapheme den einzelnen Phonemen bzw.  
Phonemeinheiten zugeordnet, wobei im vorliegenden Fall bei  
20 der Zuordnung des Graphems „c“ zu der Phonemeinheit „yu“ eine  
zu den bisherigen Zuordnungen widersprüchliche Zuordnung er-  
folgt. Diese widersprüchliche Zuordnung wird festgestellt und  
das Graphem „c“ wird mit dem vorhergehenden Graphem „c“ zu  
„cc“ zusammengefaßt. Dies führt wiederum zu einer korrekten  
25 Zuordnung der Grapheme zu den Phonemen.

Im Schritt S15 wird wiederum geprüft, ob widersprüchliche Zu-  
ordnungen erfolgt sind. Werden derartige widersprüchliche Zu-  
ordnungen festgestellt, werden sie und die weiteren Zuordnun-  
30 gen des jeweiligen Wortes gelöscht.

Mit dem Schritt S16 wird das Verfahren beendet.

Die Anzahl der im Schritt S15 ermittelten widersprüchlichen  
35 Zuordnungen ist ein Merkmal für die Qualität der Aufbereitung  
der Ausgangsdatenbank zu der durch das Verfahren erhaltenen  
Datenbank mit den einzelnen Graphem-Phonem-Zuordnungen.

die zuletzt ausgegebene Gruppierung  $Gr_1$  und das zuletzt ausgegebene Phonem  $Ph_1$  eingegeben.

5 Dieses Netzwerk wird mit den Wörtern der mit dem erfindungsgemäßen Verfahren aufbereiteten Datenbank trainiert, deren Graphem-Phonem-Zuordnungen keinen Widerspruch zu den übrigen Graphem-Phonem-Zuordnungen darstellen, das heißt, den Wörtern, deren Grapheme korrekt den Phonemen zugeordnet werden konnten.

10

Das neuronale Netzwerk ermittelt jeweils für den mittleren Buchstaben  $B_2$  ein Phonem, wobei der jeweils im Kontext vorhergehenden und nachfolgende Buchstabe und das dem zu ermittelnden Phonem vorhergehende Phonem  $Ph_1$  berücksichtigt werden. Stellen die beiden aufeinanderfolgenden Buchstaben  $B_2$  und  $B_3$  eine Graphemeinheit dar, so ergibt sich eine Ausgabe für die Gruppierung  $Gr$  von zwei. Ist der Buchstabe  $B_2$  kein Bestandteil einer aus mehreren Buchstaben bestehenden Graphemeinheit, so wird als Gruppierung  $Gr$  eine eins ausgegeben.

20

Auf der Eingangsseite wird die jeweils letzte Gruppierung  $Gr_1$  berücksichtigt, wobei im Falle einer Gruppierung von  $Gr_1$  von zwei dem mittleren Buchstaben  $B_2$  kein Phonem  $Ph$  zugeordnet wird, da dieser Buchstabe bereits mit der letzten Graphemeinheit berücksichtigt worden ist. In diesem Fall wird der zweite Buchstabe der Gruppierung übersprungen.

25

Beim Training des neuronalen Netzwerkes werden, wie es an sich bekannt ist, jeweils die Werte für die Eingangsknoten und für die Ausgangsknoten dem neuronalen Netzwerk vorgegeben, wodurch sich das neuronale Netzwerk die jeweiligen Zuordnungen im Kontext der Wörter aneignet.

30

Es kann zweckmäßig sein, mehr als drei Buchstaben an der Eingangsseite des neuronalen Netzwerkes vorzusehen, insbesondere bei Sprachen, wie der englischen Sprache, in welcher mehrere Buchstaben zur Darstellung eines einzigen Lautes verwendet

35



Mit einer solchen elektronischen Datenverarbeitungsvorrichtung 1 kann auch das erfindungsgemäße Verfahren zum Aufbereiten einer Datenbank ausgeführt werden, wobei das Verfahren wiederum in Form eines Computerprogrammes im Speicher 4 gespeichert ist und von der zentralenessoreinheit 3 zur Ausführung gebracht wird, wobei es eine Ausgangsdatenbank, die ein Wörterbuch in Schreib- und Lautschrift darstellt, in eine Datenbank aufbereitet, in der die einzelnen Laute, die Phoneme, den einzelnen Buchstaben bzw. Buchstabenkombinationen, den Graphemen zugeordnet sind.

Die Zuordnung der einzelnen Grapheme zu den einzelnen Phonemen kann in der aufbereiteten Datenbank durch Leerzeichen gespeichert werden, die zwischen den einzelnen Phonemen und Graphemen eingefügt werden.

Die das erfindungsgemäße Verfahren bzw. das neuronale Netzwerk darstellenden Computerprogramme können auch auf beliebige elektronisch lesbare Datenträger gespeichert werden und so auf eine weitere elektrische Datenverarbeitungsvorrichtung übertragen werden.

Die Erfindung ist oben anhand eines Ausführungsbeispiels beschrieben, mit dem eine Datenbank für die Sprachsynthese erzeugt wird. Im Rahmen der Erfindung ist es selbstverständlich auch möglich, die erfindungsgemäß erzeugte Datenbank bei der Spracherkennung zu verwenden, da Spracherkennungsverfahren oftmals Datenbanken mit Graphem-Phonem-Zuordnungen gebrauchen.

30

Eine Spracherkennung kann bspw. mit einem neuronalen Netzwerk ausgeführt werden, das mit der erfindungsgemäß erstellten Datenbank trainiert worden ist. An der Eingangsseite weist dieses neuronale Netzwerk vorzugsweise drei Eingangsknoten auf, an welchen das in ein Graphem umzusetzende Phonem eingegeben und falls vorhanden zumindest ein im Wort vorhergehendes und ein nachfolgendes Phonem eingegeben werden. An der Ausgangs-

**THIS PAGE BLANK (USPTO)**